DOCUMENT RESUME

ED 384 647 TM 023 860

AUTHOR Wang, Lin; And Others
TITLE Effects of Nonnormal Data on Parameter Estimates in Covariance Structure Analysis: An Empirical Study.
PUB DATE Apr 95
NOTE 30p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Behavior Patterns; Children; *Chi Square; *Error of Measurement; *Estimation (Mathematics); Evaluation Methods; Least Squares Statistics; Maximum Likelihood Statistics; *Sample Size; *Structural Equation Models
IDENTIFIERS *Covariance Structural Analysis; Kurtosis; *Nonnormal Distributions; Parameter Identification; Self Report Measures; Skew Curves; Standardization

ABSTRACT

Research in structured equation modeling (SEM) suggests that nonnormal data will invalidate chi-square tests and produce erroneous standard errors. However, much remains unknown about the extent to which, and the conditions under which nonnormal data can affect SEM application, especially when excessive skewness and kurtosis are present in data. Using different sample sizes and estimation methods, this empirical study investigates how parameter estimates, standard errors and selected fit indices are affected by nonnormal data with excessive kurtosis and skewness. The standardization data from a children's self-report behavioral assessment were used, drawing 100 random samples of 200, 500, and 1,000 each from the total population of 5,410 children. Findings indicate that: (1) normal theory maximum likelihood and generalized least squares estimators are fairly consistent and almost identical; (2) standard errors tend to underestimate the true variation of the estimators, but the problem is not very serious for a large sample (n=1,000) and conservative (99%) confidence intervals; and (3) the adjusted chi-square tests seem to be able to yield acceptable results given an appropriate sample size. One figure and seven tables present analysis results. (Contains 16 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

EFFECTS OF NONNORMAL DATA ON PARAMETER ESTIMATES IN COVARIANCE STRUCTURE ANALYSIS: AN EMPIRICAL STUDY

LIN WANG

PENNSYLVANIA POWER AND LIGHT COMPANY

VICTOR L. WILLSON

TEXAS A&M UNIVERSITY

XITAO FAN

UTAH STATE UNIVERSITY

ED 384 647

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

LIN WANG

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

PAPER PRESENTED AT THE ANNUAL MEETING OF THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, APRIL 18-22, 1995, SAN FRANCISCO, CA.

TM025860

BEST COPY AVAILABLE

Abstract

Research in structural equation modeling (SEM) suggests that nonnormal data will invalidate chi-square tests and produce erroneous standard errors. However, much remains unknown about the extent to which, and the conditions under which nonnormal data can affect SEM application, especially when excessive skewness and kurtosis are present in data. Using different sample sizes and estimation methods, this empirical study investigates how parameter estimates, standard errors and selected fit indices are affected by nonnormal data with excessive kurtosis and skewness. Our findings include (1) normal theory maximum likelihood (ML) and generalized least squares (GLS) estimators are fairly consistent and almost identical; (2) standard errors tend to underestimate true variation of the estimators, but the problem is not very serious for large sample (n = 1,000) and conservative (99 percent) confidence intervals; (3) the adjusted chi-square tests seem to be able to yield acceptable results given an approriate sample size.

Effects of Nonnormal Data on Parameter Estimates in Covariance Structure Analysis: An Empirical Study

Introduction

Structural equation modeling (SEM) is a statistical technique that has now been used quite often by researchers in education, psychology, and other social sciences (Johnson & Wichern, 1992; Joreskog & Sorbom, 1981). In particular, SEM, "and its important special cases of covariance structure analysis and confirmatory factor analysis, has become an important tool for testing theories with nonexperimental data" (Bentler, 1994, p. 237). The technique is bridging researchers' substantive thinking and the way of doing data analysis (Bollen, 1989). Normal theory maximum likelihood (ML) and generalized least squares (GLS) are the two typical estimation methods that are available in computer programs for SEM. Both ML and GLS require multinormality in the data for the computational results to be valid. Absence of multinormality (normality for short) alone can invalidate standard errors and chi-square test statistics. In the real world of research, however, SEM has often been applied to data that lack the evidence of normal distributions (Bentler, 1994; Micceri, 1989). In other words, use of nonnormal data in SEM application is by no means a rarity, if not a common practice.

Concern with the possible misleading results of SEM application associated with nonnormal data has led to quite extensive research in this area. Estimation procedures and test statistics that are less sensitive to or correct for nonnormality in SEM application have been proposed, such as the arbitrary distribution function (ADF) procedure (Brownie, 1984), scaled test statistics (Chou, Bentler & Satorra, 1991), weighted least squares and

elliptical estimators (Bollen, 1989), to name a few. These alternative estimation procedures or new test statistics can be complicated and are usually available only in some specialized computer programs. Another important line of research has focused on the asymptotic robustness of normal theory methods. From the viewpoint of applied researchers, this line of research can provide more important information about appropriate use of SEM than the aforementioned alternative estimation procedures or new test statistics.

Bentler (1994) has pointed out that "asymptotic robustness theory promises to extend the range of applicability of the computationally simpler ML and GLS estimators to situations where the more difficult distribution-free methods might seem to be needed" (p. 240). Generally speaking, the study of the asymptotic robustness indicates that parameter estimates can be consistent with nonnormal data (Bollen, 1989), but standard errors of the parameter estimates and test statistics are questionable, depending on the characteristics of nonnormality in data. In a Monte Carlo study by Chou et al (1991), the simulation data were manipulated to assume different degrees of skewness and kurtosis to reflect a variety of the characteristics of nonnormality that one might expect in research. Their results showed little difference among the estimates of a parameter under different nonnormality conditions. It was also found that ML test statistics and standard errors appeared "to be quite robust to nonnormality when data had either symmetric and platykurtic distributions, or nonsymmetric and zero kurtotic distributions" (Chou et al, 1991, p. 347). In a similar study of confirmatory factor analysis (Hu, Bentler, & Kano, 1992), it was found that, when both common factors and unique factors were nonnormally distributed, ML and

GLS methods outperformed ADF method at all but the largest sample size (n = 5,000) in terms of mean test statistic across 200 replications and the rate of rejecting the true model. Between ML and GLS methods, the effect of sample size was evident on mean test statistics, with GLS giving better results than ML for smaller sample sizes.

Although the literature of robustness study has added to the understanding of the effect of nonnormality on parameter estimates, test statistics and standard errors, much remains to be learned about the asymptotic robustness theory (Bentler, 1994). First, parameter estimates are known to be asymptotically consistent even with nonnormal data when the sample size is large. It is, however, not clear whether, or how well, this holds across all situations where excessive kurtosis and skewness are both present in the data. Second, the usefulness of standard errors in SEM with nonnormal data has not been fully studied. A standard error of a parameter estimate is used in constructing a confidence interval based on a desired probability level. The extent to which the confidence intervals calculated from an estimate and its standard error cover the corresponding population value is a criterion for the usefulness and the quality of the standard error. Very little, if any, study of this nature has been reported. Third, test statistics, like chi-square tests, have received most of the attention in SEM with nonnormal data. However, in a simulation type of study, use of mean chi-square test statistics across replication samples as a criterion for the quality of test statistics may not be appropriate when nonnormal data are involved. This is because the sampling distribution of the chi-square test statistics is unknown when the normality assumption is not satisfied in SEM; the empirical distribution of chi-square test statistics across replications is not normal. Therefore, mean values of

such test statistics may not be a good descriptive measure, and median values may be preferred, instead. Fourth, the performance of fit indices other than chi-square test statistic has not been studied when nonnormal data are used. Bentler & Bonett (1980) and Bollen (1989) suggested that other fit indices should always be reported along with chi-square test results. Chi-square test statistics should not be used as the sole criterion for model fit, as the chi-square test is easily influenced by data characteristics like normality and by sample size. Fifth, empirical studies of nonnormality in SEM applications have mainly been done with simulation data whose characteristics are artificially predefined. These simulation studies are important and helpful in many respects. Nevertheless, findings from an empirical study using real data will certainly add to the understanding of the topic and have more practical significance.

The goal of this study is to contribute to the existing body of knowledge of the effect of nonnormal data in SEM application in general. In particular, this study is intended to explore empirically the effect of nonnormal data with excessive skewness and kurtosis on parameter estimates, including standard errors, as well as on model fit indices. The objectives of this study are to (a). use a real data set as the population, resample from this population to simulate data collection in real research situations, and then fit a simple structural equation model to both the population and the resampled data; (b) examine parameter estimates and standard errors under different conditions defined by estimation methods (ML and GLS) and sample sizes, with the quality of standard errors being judged by the percentage of confidence intervals that may cover the population parameter values; (c) evaluate chi-square tests and adjusted chi-square tests with regard to the rates of their

not rejecting the true model; (d) evaluate the performance of several other fit indices that are either normed or non-normed.

Methods

Data

The population data for this empirical study were the standardization data of children's self-report from a newly developed behavioral assessment, Behavioral Assessment System for Children (BASC) (Reynolds & Kamphaus, 1992). This instrument is a personality inventory of twelve scales with statements to be responded to as True or False. Raw scores were linearly transformed, but not normalized, to T scores for all the scales by the BASC developers. Four scales were selected as the variables of interest for this study: (1) Relations with parents, (2) Interpersonal relations, (3) Self-esteem, and (4) Depression. Data from 5,410 subjects were available and used as the population data, from which random samples of different sizes were taken and analyzed. All four variables have nonnormal distributions that are characterized by either excessive kurtosis or excessive skewness, or both. The summary statistics of the four variables are in Table 1 below.

Insert Table 1 about here

---

100 random samples of size 200, 500 and 1,000 each were drawn from the population (N=5,410) by means of sampling without replacement. This was done using

the SAS uniform random number generator (SAS Inc., 1990) to simulate the practice of taking simple random samples from a population. Values of kurtosis and skewness of the four variables were obtained from each of the 300 samples. The means, medians, minimum and maximum values of the distributions of the kurtosis and skewness are given in Table 2. These values demonstrate the excessiveness of kurtosis and skewness in the sample data, and therefore the degree of nonnormality of the sample data. Although the sample mean values of kurtosis and skewness are close to the population values in Table 1 and many of them are already excessive, the sample minimum and maximum values of kurtosis and skewness are more extreme. Estimations from the data with such extreme kurtosis and skewness will provide more information about to what extent nonnormality may impact parameter estimation and evaluation of fit indices in structural equation modeling.

Insert Table 2 about here

Model and analysis

A theoretical model about the relationship among the four variables was hypothesized. Three variables, i.e. Relations with parents (X1), Self-esteem (X2), and Interpersonal relations (X3), were taken as the indicators of a latent variable (F1), Personal adjustment, according to the test developer (Reynolds & Kamphaus, 1992). Depression (Y) was the only one indicator for the other latent variable (F2), Clinical

adjustment. Personal adjustment decrement was believed to lead to clinical maladjustment. The path diagram of the model is in Figure 1 and the numbers are the population parameter values.

Insert Figure 1 about here

In the application of structural equation modeling in research, it is generally acknowledged that substantive knowledge plays a crucial role in model construction, model fitting and evaluation, and interpretation of results (Bollen, 1989; Joreskog, 1993; Pedhazur & Schmelkin, 1991). The purpose and nature of this study dictated that the emphasis of this study was not on the substantive interest in the proposed behavioral model in Figure 1. The model was used for the sole purpose of facilitating the study of the nonnormality issue.. Also, the size of the model was restricted to four variables in order to focus the study on the topic of central interest and to avoid the unnecessary distraction that a complex model might cause. A variety of fit indices are usually available for model evaluation. A detailed discussion was given by Tanaka (1993) of the fit indices regarding their differences along six conceptual dimensions. In this study, nine types of fit indices were chosen for evaluation and they are goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), p-value of chi-square test (P-CHI), p-value of adjusted chi-square test (P-ACHI), Bentler's comprehensive fit index (CFI), Bentler and Bonett's normed fit index (BNOR) and non-normed fit index (BNON), Bollen's normed RHO1

index (RHO1) and non-normed DELTA2 (DELTA2). The normed indices range between 0 and 1; the non-normed indices can have values slightly over 1.

PROC CALIS procedure in SAS program (version 6.08 for Windows) was used in all the analyses. Although popular computer programs are available that are specifically used for SEM, like LISREL and EQS, as a comprehensive and powerful statistical package, SAS provides more programming tools suitable for simulation type study such as this one. Besides, PROC CALIS under SAS produces more fit indices than either LISREL or EQS. Both maximum likelihood (ML) and generalized least squares (GLS) options are available in PROC CALIS and were used in all the cases. The fitting functions in PROC CALIS (SAS, 1990a, pp. 292-293) are:

$$F = Tr(\mathbf{SC}^{-1}) - \underline{n} + \log(\det(\mathbf{C})) - \log(\det(\mathbf{S}))$$

for ML method, and

$$F = 0.5\ Tr(\mathbf{S}^{-1}(\mathbf{S} - \mathbf{C}))^2$$

for GLS method, where, $\underline{n}$ is the number of manifest variables, **C** is the predicted covariance matrix, and **S** is the observed sample covariance matrix. The model was defined in LINEQS format, one of the four model specification formats under PROC CALIS, and the sample covariance matrix was analyzed. The model was first fitted to the population data and all the population parameters were obtained. The same model was then fitted to all the random sample data. The macro processing in SAS was applied in automating the whole process of taking 300 random samples from the population data, running the PROC CALIS procedure for all the 300 samples, and outputting the selected results for further analysis.

Results

Parameter estimates and standard errors

Eight free parameters were obtained for the population data. The same parameters were also estimated from the 300 random samples, i.e. 100 replications for each of the three sample size conditions (n = 200, n = 500, n = 1,000). With eight estimated parameters, three sample sizes, and two estimation methods, a total of 48 distributions of parameter estimates was obtained. The univariate normality test using the SAS procedure PROC UNIVARIATE NORMAL (SAS Inc., 1990b, p. 627) indicated that 14 distributions were not normal ($P < .05$). Of the 14 distributions, 10 were from the sample size of 200. The summary statistics of the 48 distributions are given in Table 3. Although most of the distributions are normal and means and standard deviations are sufficient to describe such distributions, medians, minimum and maximum values of the distributions are also included to demonstrate the degree of variability in the distribution of a particular parameter estimate. Because the parameter estimates were computed from nonnormal data that had excessive skewness or kurtosis, the qualities of these estimates are not clear. As an empirical study, additional information about the distributions of these parameter estimates can help to gain more knowledge of the characteristics of parameter estimation in SEM with nonnormal data.

It is apparent that results from ML and GLS are practically identical except for some trivial differences. Sample size plays a noticeable role in estimation. The mean and median parameter estimates for n = 1,000 are the closest to the population values at the bottom of Table 3. These estimates also have the smallest standard deviations and the

smallest differences between maximum and minimum values. With $\underline{n}$ = 200, the mean estimates are more different from the population values and greater variation is observed in the distributions of the estimates. The estimates from the sample size of 500 are better than those from the sample size of 200 but not as good as those from the sample size of 1,000.

Insert Table 3 about here

The standard errors for the eight parameter estimates were also collected. To examine the usefulness of the standard errors, confidence intervals were constructed at two conventional levels of 95 and 99 percent, for each parameter estimate. Percentages of these confidence intervals that covered the respective population parameters were calculated from the frequencies based on 100 replications. The results in Table 4 show that there is little difference between the estimation methods. Under the ML method in Table 4, of the percentages for the nominal 95 percent confidence intervals, one ($\lambda 2$) is exactly 95 percent. Another three are also fairly close: 93 for γ, 93 for ζ ($\underline{n}$ =1,000) and 92 for ζ ($\underline{n}$ = 500). The lowest percentage of coverage is 72 for $\delta 1$. A similar pattern of the distribution of the percentages can be found for the 99 percent confidence intervals. Three values are at or even above the desired 99 level. Another 10 percentages are at or greater than 95. The number of confidence intervals that cover the population values is the greatest when the sample size is 1,000 and the smallest when the sample size is 200.

With the largest sample size (n = 1,000) and across the eight parameters, the median coverage rate of the population values is 97 percent, ranging from 91 to 100 percent, for the nominal 99 percent confidence intervals; the median coverage rate is 88.5 percent, ranging from 72 to 95 percent, for the nominal 95 percent confidence intervals.

Insert Table 4 about here

Fit indices

The summary statistics of the nine fit indices across the 100 replications were listed in Table 5. Because the univariate distributions of these fit indices were highly skewed, median values are used in place of means. Minimum and maximum values are also given for reference. The medians of the 9 indices estimated with GLS and ML methods are identical to the second decimal place in most cases. Difference due to sample size is not noticeable among the medians of GFI, AGFI, BNOR BNON, RHO1 and DELTA2. The influence of sample size is reflected, however, in the minimum values of these indices. For both P-CHI and P-ACHI values, differences due to sample size are expectedly evident.

Insert Table 5 about here

An attempt was made to explore the performance of chi-square tests (overall and adjusted) in the form of nonrejection rates at the conventional .05 probability level. In other words, the number of times that the chi-square tests did not reject the null hypothesis has been counted. The results in Table 6 depict the percentages of nonrejection by the chi-square and the adjusted chi-square tests for the three sample size conditions. The adjusted chi-square tests made fewer rejections than overall chi-square tests across all the three sample size conditions. For both the chi-square and the adjusted chi-square tests, the lowest rejection rates were registered with the sample size of 500 (only four percent for the adjusted chi-square tests and 18 percent for the chi-square tests), while the highest rejection rates are seen in the tests with the largest sample sizes, i.e. $\underline{n}$ =1,000, (22 and 47 percent for the adjusted chi-square and the chi-square tests, respectively).

Insert Table 6 about here

In Table 7 are the coefficients of variation (CV) of seven fit indices other than P-CHI and P-ACHI. Coefficients of variation is calculated by dividing the mean of a distribution with its standard deviation and then multiplying 100. CV is a unitless measure and can be used to compare the variability of distributions that have different standard deviations and different means. The coefficients of variation of the distributions of the fit indices based on 100 replications can serve to indicate which fit indices are less variable

given the same conditions (mainly sample size) and are therefore relatively more useful. Again, little difference is found between ML and GLS results. Therefore, only GLS results are presented in Table 7 for illustration. It appears that three indices, CFI, BNOR and DELTA2, show the smallest differences among the three sample sizes; this is especially evident between the sample sizes of 500 and 1,000.

Insert Table 7 about here

Discussion

Parameter estimation

ML and GLS parameter estimates in covariance analysis are said to be consistent even with nonnormal data (Bollen, 1989). Also, when the weight matrix **W** is **S** ($\mathbf{W}^{-1} = \mathbf{S}^{-1}$) in the GLS fitting function, the GLS estimator is asymptotically equivalent to the ML estimator (Bentler & Weeks, 1980). The results in Table 3 show that this equivalence between ML and GLS estimators seems to also hold with nonnormal data. Compared with the population parameters, the mean estimates across 100 replications approach the population values as sample sizes increase from 200 through 500 to 1,000. The differences between the minimum and maximum estimates decrease noticeably with the increase of sample sizes. It seems therefore that the quality of parameter estimates is not much of concern even with nonnormal data, provided adequately large samples are used. It is hard to define how large a sample should be. From what is given in Table 3, it is

found that, even for such a small and simple model with only eight free parameters, the estimates from $\underline{n} = 200$ are rather unsatisfactory. In fact, the estimates appear to start showing a sign of stabilizing when the sample size reaches 500. Between the ML and GLS results of ML and GLS, differences are small or none to the second decimal place. A certain pattern in the results, however, does appear to exist between the two methods. Practically no difference is found in the values of the loading, $\lambda 1$, $\lambda 2$, γ. On the other hand, ML estimates of the error (δ) and disturbance (ζ) are all slightly higher than those of GLS estimates. Considering the sizes of the δs and ζ, the differences are really trivial.

The results in Table 4 can be very helpful in understanding the usefulness of standard errors of parameter estimates in covariance analysis with nonnormal data. Several observations can be made regarding the results. First, differences between ML and GLS results are very small and exhibit no apparent pattern across either sample sizes or parameters. Second, the results in Table 4 indicate that the theoretical standard errors given for estimators tend to underestimate the true variation of the estimators, since the percentages of both 95 and 99 percent confidence intervals that contain the population parameters are generally lower, or even much lower than the nominal levels. The smaller (underestimated) theoretical standard errors result in narrower confidence intervals that cover fewer parameters than expected for the specified .05 and .01 probability level. Third, at both 95 and 99 percent confidence interval levels, the estimates of error variance (δ) yield fewer confidence intervals that cover the parameter values than do other estimates. In other words, the confidence intervals for error variance estimates (δ) might be less useful than the confidence intervals for other estimates. Fourth, large sample size

as well as the conservative .01 probability level appears to produce standard errors that are closer to the true variation of the parameter estimators. The empirical rate of the confidence intervals containing the parameters is close to the corresponding nominal level. Judging by an actual median value of 97 percent for the nominal 99 percent confidence interval with $\underline{n}$ = 1,000, the impact of nonnormality with excessive kurtosis and skewness does not seem to be very serious, and the standard errors might be more useful than has been portrayed in the literature. This is at least so for the conservative 99 percent confidence intervals in this investigation. However, the problem is that the performance of standard errors may depend on the type of parameters, i.e., δ, γ, ζ, or ϕ, and, possibly, on model complexity. Also, only 100 replications were used in this study and the largest sample size is limited to 1,000. It can be reasoned that a study with more replications and larger sample sizes will provide more reliable and convincing findings in this respect.

Fit indices

The results depicted in Table 6 are noteworthy. The percentages of the chi-square tests and the adjusted chi-square tests that do not reject the null hypothesis are rather different, with the latter being much closer to the nominal probability level (95 percent). These results are from GLS method and are only about 1 to 3 percent higher than ML results. The adjusted chi-square tests perform much better than the chi-square tests across the three sample sizes. For the adjusted chi-square tests with the sample size of 500, 96 percent of the tests has not rejected the null model in comparison with the expected 95 percent. It is not clear why the adjusted chi-square tests have done so well at this sample size. In general, however, that the adjusted chi-square tests outperform the chi-square

tests could be due to the correction made in the former. In PROC CALIS (SAS, 1990, p. 305), the adjusted chi-square test is the chi-square test corrected for elliptical distribution, a distribution that is symmetrical but is not normal due to multivariate kurtosis. It is possible that, having at least adjusted for the kurtosis problem in the data, the adjusted chi-square tests are able to yield better results than the chi-square tests which require multinormality for the data. Therefore, the adjusted chi-square tests, instead of the conventional chi-square tests, might be recommended for model fit evaluation when data have a variety of excessive kurtosis and skewness.

Within the results of the adjusted chi-square tests, however, the differences due to sample size conditions call for attention. The nonrejection rate of the tests with a sample size of 500 is the best and ideal. The tests with the largest sample size ($\underline{n}$ = 1,000) make too many rejections (22 percent) to be useful. This high rejection rate might have been due to the high statistical power that results from a large sample size and a simple model with a very limited number of parameters to be estimated. On the other hand, with a sample size of 200, 11 percent of the tests have rejected the null model, a rate much higher than that for the sample size condition of 500. This seems to contradict the rule that a higher rejection rate tends to be associated with greater statistical power that is often the product of large sample size. However, a review of the distributions of the sample kurtosis and skewness across 100 replications in Table 2 finds greater variations of sample kurtosis and skewness for the sample size condition of 200 than for the sample size condition of 500. And possibly, under the sample size condition of 200, it is the greater variations of sample kurtosis and skewness that have made the correction for elliptical

distribution less effective in the adjusted chi-square tests, leading to a higher rejection rate of the tests.

Variability of a fit index across replications might be used as another practical criterion in evaluation of the usefulness of the fit index. Where little is known about the distribution characteristics of a fit index, as is the case for many fit indices in SEM, a researcher might want to use the fit index that is resistant to sampling error or the influence of some other factors, say, estimation method. The p-values of the adjusted chi-square tests (P-ACHI) show relatively smaller variation than the p-values of the chi-square tests (P-CHI). Like the higher nonrejection rates of the adjusted chi-square tests discussed earlier, the smaller variation in the probability values of the adjusted chi-square tests make this type of test preferable to the chi-square tests for the kind of nonnormal data used in this study. Of the other seven fit indices in Table 7, the coefficients of variation for CFI, BNOR, and DELTA2 appear to be affected relatively little by sample size. The others show greater variability, especially at the sample size of 200. The adjusted goodness-of-fit index (AGFI) seems to be the poorest. In Tanaka's typology of fit indices (1993), BNOR and CFI are classified as sample size dependent since sample size directly enters the calculation of these indices; DELTA2 (called incremental fit index, or IFI) is not sample size dependent. In Table 7, however, BNOR, CFI, and DELTA2 are almost the same in magnitude of variation. It is not clear why such results seem to differ from Tanaka's typology. Since the coefficients of variation in Table 7 are calculated from the actual empirical distributions of fit indices across 100 replications, these results may suggest that the effect of sample size on BNOR, CFI, and DELTA2 can be trivial.

Drawing upon the findings from Table 7, two suggestions are in order regarding the selection of fit indices in studies that involve nonnormal data of the type in this study. First, be aware of the difference between the conventional chi-square tests and the adjusted chi-square tests and their respective probability values; the adjusted chi-square tests should be used. Second, when concern with the effect of sample size is at issue, BNOR, CFI and DELTA2 are better fit indices to use than Bentler's non-normed index (BNON), Bollen's RHO1, GFI and AGFI.

Conclusion

A relatively simple structural equation model has been fitted to the nonnormal data in 100 replications for each of three sample sizes; the effect of nonnormal data on parameter estimation and fit indices of model evaluation has been investigated in this empirical study. Because the 300 replication samples were randomly drawn from a defined population of real data, the findings of this study have realistic and practical relevance to researchers who need to use structural equation modeling technique but are concerned over the problem of nonnormal data.

This study has made the following contributions to the study of covariance structure analysis for nonnormal data that have excessive kurtosis and skewness. First, this study has empirically investigated the behavior and the usefulness of standard errors of parameter estimates by looking at the actual percentages of confidence intervals that cover the population values under the conditions defined by estimation methods and sample sizes. Second, this study has found that the adjusted chi-square test that corrects for elliptical distribution can yield acceptable results even for nonnormal data, given an

appropriate sample size that balances the statistical power of the test against the sampling variation. Third, in selecting fit indices whose values lie roughly between 0 and 1, Bentler's comparative fit index, Bentler & Bonett's normed index, and Bollen's DELTA2 appear to be more stable across sample sizes and should be used when sample size is an issue or a concern in analysis. In particular, the findings here will provide a general reference about what a researcher might expect from doing structural equation modeling with nonnormal data in terms of parameter estimates, standard errors and fit indices for model evaluation.

On the other hand, it is also important to realize the limitation of this study with regard to the implication of the findings and applicability to other situations. Although the kurtosis and skewness characteristics of the population and sample data should be fairly representative of nonnormal data encountered in behavioral research, both the population and the samples are limited in size and variation. The number of replications is also not sufficiently large as 100 replications are usually the minimum number required in a simulation type of empirical study. Another limitation is the lack of comparison among the fit index values. Because the fitted model is almost saturated (df = 2 only), the fit is bound to be good. However, potential differences among the different fit indices may have been made less visible. The findings of this study should not be taken as definitive answers to the problems involved and are subject to verification from future studies with larger data sets , more replication samples, and different models.

References

Bentler, P. M. (1994). On the quality of test statistics in covariance structure analysis: Caveat emptor. In C. R. Reynolds (Ed.), Cognitive assessment: A multidisciplinary perspective (pp. 237-260). New York: Plenum Press.

Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bentler, P. M. & Weeks, D. G. (1980). Linear structural equations with latent variables. Psychometrika, 45, 289-308.

Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

Brownie, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 37, 62-83.

Chou, C. P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis: A monte carlo study. British Journal of Mathematical and Statistical Psychology, 44, 347-357.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? Psychological Bulletin, 112, 351-362.

Johnson, R. A. & Wichern, D. W. (1992). Applied multivariate statistical analysis (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Joreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.) Testing structural equation models (pp. 294-316). Newburry, CA: Sage.

Joreskog, K. G. and Sorbom, D. (1981). LISREL V: Analysis of linear structural relationships by the method of maximum likelihood. Uppsala: University of Uppsala.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin, 105, 156-166.

Pedhazur, E. J. & Schmelkin, L. P. (1991). Measurement, design, and analysis: An integrated approach (Student edition). Hillsdale, NJ: LEA.

Reynolds, C. R. & Kamphaus, R. W. (1992). BASC Self-Report: SRC-C (Ages 8 - 11). Circle Pines, MN: AGS.

SAS Institute Inc. (1990a). SAS/STAT user's guide: Volume 1, ANOVA-FREQ (Version 6, 4th ed.). Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1990b). SAS procedures guide (Version 6, 3rd ed.). Cary, NC: SAS Institute Inc.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.) Testing structural equation models (pp. 10-39). Newburry, CA: Sage.

Table 1. Summary Statistics of the Four Variables for the population (N=5,410).

Variable	M	S	Skewness	Kurtosis
X1 (Relation with parent)	50.12	9.75	-1.85	3.46
X2 (Interpersonal relation)	50.16	10.23	-1.59	2.02
X3 (Self-esteem)	49.83	9.90	-1.36	0.79
Y (Depression)	49.97	10.01	-1.17	0.48

Table 2. Sample Distributions of Skewness and Kurtosis.

	Sample Size					
	200		500		1,000	
Variable	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
X1	-1.84	3.48	-1.84	3.38	-1.84	3.37
	-2.49	0.43	-2.13	1.81	-2.01	2.12
	-1.20	6.67	-1.52	5.19	-1.59	4.24
X2	-1.62	2.20	-1.59	2.06	-1.57	1.95
	-2.07	0.82	-1.88	0.94	-1.78	1.33
	-1.29	4.54	-1.28	3.49	-1.42	2.75
X3	-1.40	0.98	-1.38	0.86	-1.36	0.78
	-1.75	-0.17	-1.63	0.16	-1.51	0.28
	-1.00	2.39	-1.17	1.75	-1.20	1.32
Y	-1.20	0.61	-1.18	0.52	-1.16	0.47
	-1.65	-0.19	-1.41	-0.24	-1.34	0.07
	-0.93	2.18	-0.93	1.33	-1.03	1.15

Note. The three entries in each cell are: mean, minimum, and maximum values from 100 samples.

Table 3. Parameters and Parameter Estimates for the Three Sample Size Conditions

Method	λ2	λ3	γ	ϕ	δ1	δ2	δ3	ζ
GLS	1.27	1.14	1.50	37.70	52.40	45.12	49.05	20.75
(n = 200)	.21	.19	.26	12.17	8.69	8.60	7.90	6.59
	1.24	1.11	1.46	36.78	52.93	44.35	47.24	20.88
	.89	.79	1.02	16.19	29.74	26.98	34.41	0.00
	1.79	1.60	2.43	84.11	69.73	64.81	67.76	34.79
ML	1.27	1.14	1.51	37.39	54.25	46.35	50.62	20.85
(n = 200)	.22	.19	.27	12.20	9.02	8.68	7.94	6.91
	1.24	1.11	1.47	36.29	54.79	45.31	48.75	21.02
	.89	.79	1.02	16.08	32.16	27.05	35.53	0.00
	1.81	1.60	2.45	83.88	73.38	65.77	68.29	35.02
GLS	1.23	1.10	1.43	39.59	56.41	46.07	51.23	21.01
(n=500)	.13	.10	.14	7.60	5.35	4.91	5.21	3.80
	1.22	1.10	1.40	39.27	56.12	46.13	51.01	20.84
	1.00	.89	1.14	21.42	44.67	32.06	40.54	12.10
	1.58	1.35	1.88	62.64	67.80	59.53	63.23	30.33
ML	1.23	1.10	1.43	39.47	57.13	46.50	51.86	21.07
(n=500)	.13	.10	.14	7.60	5.56	4.87	5.21	3.86
	1.22	1.10	1.40	39.25	57.02	46.85	51.56	20.88
	1.00	.87	1.14	21.25	45.04	32.48	41.03	12.12
	1.59	1.35	1.88	62.64	71.43	59.72	63.74	30.37
GLS	1.23	1.08	1.42	39.91	56.44	46.82	52.06	21.46
(n=1,000)	.08	.06	.08	4.35	4.59	3.44	3.85	2.48
	1.22	1.08	1.41	39.92	56.29	47.29	51.82	21.44
	1.07	.97	1.27	28.72	45.22	37.56	41.03	16.01
	1.41	1.28	1.68	49.44	68.65	56.16	60.62	28.25
ML	1.23	1.08	1.42	39.79	57.10	47.18	52.65	21.50
(n=1,000)	.08	.06	.08	4.34	4.68	3.47	3.82	2.54
	1.23	1.08	1.41	39.79	56.82	47.50	52.48	21.48
	1.07	.97	1.27	28.64	45.60	37.65	41.86	15.73
	1.41	1.28	1.69	49.10	69.81	56.85	61.13	28.43
GLS	1.22	1.08	1.43	38.58	56.09	46.72	52.41	21.58
ML	1.23	1.08	1.43	38.48	56.56	46.89	52.83	21.56

Note. The five entries in each cell are: mean, standard deviation, median, minimum, and maximum values obtained from 100 replication samples. The last two rows at the bottom of the table are population parameter values (N = 5,410).

Table 4. Percentages of Confidence Intervals (C.I.) of the Parameter estimates from 100 Replications That Cover the Population Parameters.

	Sample Size					
Parameter	200	500	1000	200	500	1000
		GLS			ML	
(95% C.I.)						
λ2	86	85	89	83	85	89
λ3	88	89	93	91	89	95
γ	82	85	93	82	85	93
ϕ	79	83	88	79	84	88
δ1	78	84	74	78	85	72
δ2	80	88	90	80	87	87
δ3	84	80	86	86	83	85
ζ	92	92	93	90	92	93
		GLS			ML	
(99% C.I.)						
λ2	95	91	98	95	91	96
λ3	97	98	99	97	99	100
γ	93	94	98	93	94	98
ϕ	86	92	98	85	92	98
δ1	83	95	89	83	96	91
δ2	91	94	95	93	96	96
δ3	91	92	93	91	94	94
ζ	96	98	99	96	98	99

Table 5. Fit indices from fitting the model to 100 replications of three sample sizes.

	Sample Size					
Fit Index	200	500	1,000	200	500	1,000
		GLS			ML	
Goodness-of-fit	.99	1.00	1.00	.99	1.00	1.00
(GFI)	.96	.98	.99	.95	.97	.99
	1.00	1.00	1.00	1.00	1.00	1.00
Adjusted GFI	.97	.98	.99	.97	.98	.99
(AGFI)	.79	.88	.95	.75	.87	.94
	1.00	1.00	1.00	1.00	1.00	1.00
Probability > χ^2	.27	.18	.06	.28	.19	.06
(P-CHI)	.00	.00	.00	.00	.00	.00
	.98	.96	.96	.98	.96	.97
Bentler's Comparative	1.00	1.00	1.00	1.00	1.00	1.00
Fit Index (CFI)	.98	.99	1.00	.93	.97	.99
	1.00	1.00	1.00	1.00	1.00	1.00
		GLS			ML	
Probability > Adj. χ^2	.44	.35	.16	.43	.33	.15
(P-ACHI)	.01	.00	.00	.00	.00	.00
	.99	.97	.98	.99	.97	.98
Bentler & Bonett's	1.00	1.00	1.00	.99	.99	.99
Nonnormed (BNON)	.93	.97	.99	.80	.90	.96
	1.01	1.00	1.00	1.02	1.01	1.00
Bentler & Bonett's	1.00	1.00	1.00	.99	1.00	1.00
Normed (BNOR)	.97	.99	1.00	.93	.97	.99
	1.00	1.00	1.00	1.00	1.00	1.00
Bollen Normed Index	.99	.99	.99	.97	.99	.99
RHO1 (RHO1)	.92	.97	.99	.78	.90	.96
	1.00	1.00	1.00	1.00	1.00	1.00
Bollen Nonnormed	1.00	1.00	1.00	1.00	1.00	1.00
DELTA2 (DELTA2)	.98	.99	1.00	.93	.97	.99
	1.00	1.00	1.00	1.01	1.00	1.00

Note. The three entries in each cell are: median, minimum, maximum values. Numbers like .999 or .995 are rounded to 1.00 for practical purpose.

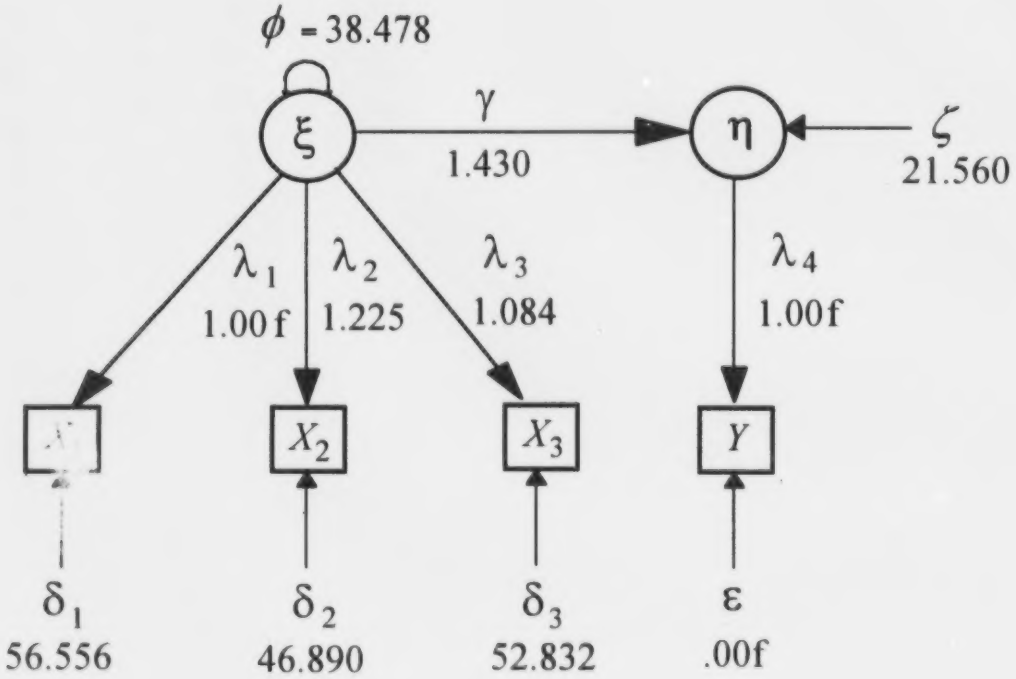
Table 6. Nonrejection Rates (P > .05) of the Chi-Square and the Adjusted Chi-Square Tests across 100 Replications.

	Sample Size		
Test	200	500	1,000
Chi-Square	75	82	53
Adjusted Chi-Square	89	96	78

Table 7. Coefficients of Covariation (CV) for the Seven Fit Indices across 100 Replications.

	Sample Size			
Fit Index	200	500	1,000	Range
GFI	1.00	.34	.22	.78
AGFI	5.22	1.72	1.14	4.08
CFI	.46	.15	.11	.35
BNON	1.50	.52	.34	1.16
BNOR	.50	.18	.11	.39
RHO1	1.53	.53	.33	1.20
DELTA2	.49	.17	.11	.38

Figure 1. The Path Diagram of the Model and the Population Parameters (N=5,410).



f = fixed value